

# The effects of halo alignment and shape on the clustering of galaxies

Marcel P. van Daalen<sup>1,2\*</sup>, Raul E. Angulo<sup>1</sup> and Simon D. M. White<sup>1</sup>

<sup>1</sup>*Max Planck Institute for Astrophysics, Karl-Schwarzschild Straße 1, 85741 Garching, Germany*

<sup>2</sup>*Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*

Accepted 2012 May 31. Received 2012 May 21; in original form 2011 October 21

## ABSTRACT

We investigate the effects of halo shape and its alignment with larger scale structure on the galaxy correlation function. We base our analysis on the galaxy formation models of Guo et al., run on the Millennium Simulations. We quantify the importance of these effects by randomizing the angular positions of satellite galaxies within haloes, either coherently or individually, while keeping the distance to their respective central galaxies fixed. We find that the effect of disrupting the alignment with larger scale structure is a  $\sim 2$  per cent decrease in the galaxy correlation function around  $r \approx 1.8 h^{-1}$  Mpc. We find that sphericalizing the ellipsoidal distributions of galaxies within haloes decreases the correlation function by up to 20 per cent for  $r \lesssim 1 h^{-1}$  Mpc and increases it slightly at somewhat larger radii. Similar results apply to power spectra and redshift-space correlation functions. Models based on the Halo Occupation Distribution, which place galaxies spherically within haloes according to a mean radial profile, will therefore significantly underestimate the clustering on sub-Mpc scales. In addition, we find that halo assembly bias, in particular the dependence of clustering on halo shape, propagates to the clustering of galaxies. We predict that this aspect of assembly bias should be observable through the use of extensive group catalogs.

**Key words:** cosmology: theory – cosmology: large-scale structure of Universe

## 1 INTRODUCTION

Investigating how matter is organized in our Universe is one of the key ways in which we can test the validity of cosmological models and constrain their parameters. By comparing theoretical predictions to observed measures of structure, such as the galaxy correlation function or the matter power spectrum, one can reject some models and fine-tune others. It is, however, important to keep in mind the limitations of theoretical models, both numerical and analytical, when making this comparison, as these may limit the applicability of the results.

There are various ways in which one can predict the organization, or “clustering”, of matter and galaxies theoretically. One can use fully hydrodynamical simulations, in which dark matter, gas and stars are treated explicitly, to follow the formation and evolution both of dark matter haloes and of the galaxies within them. For a recent review of the numerical methods behind such simulations, see Springel (2010). Such models are computationally expensive, limited to small volumes in comparison to recent galaxy surveys,

and sensitive to the *ad hoc* subgrid recipes required to include critical processes like star formation and feedback.

An alternative, first implemented by Kauffmann et al. (1999) (see also Springel et al. 2001, 2005), is to combine N-body simulations of the growth of dark matter structures with semi-analytic models of galaxy formation (e.g. White & Frenk 1991, Kauffmann, White & Guiderdoni 1993, Cole et al. 1994; see Baugh 2006 for a review). A great advantage of semi-analytic simulations is that they require comparatively little CPU time even for a large underlying N-body simulation. This allows them to be run many times and on many haloes, so that one can explore the physical processes and the associated parameters that are required to produce galaxy populations in agreement with selected observational data (such as the galaxy stellar mass, luminosity or correlation functions). Such semi-analytic simulations do not focus on the properties of individual objects, but rather on the underlying statistical properties of the entire population. In this way, the relative importance of different physical processes can be examined as a function of the time and place where they are occurring. A disadvantage of such simulations is that they provide only very crude information on the structure of individual objects.

\* E-mail: daalen@mpa-garching.mpg.de

Yet another alternative is to take the statistical approach one step further. If one is interested only in the present-day clustering of galaxies, the physical processes associated with their formation and evolution may not be relevant. One can then populate the haloes in an N-body simulation with galaxies using a purely statistical model that depends on current halo properties, for example halo mass. Galaxy clustering can then be described in terms of the clustering of the haloes. This approach is known as halo occupation distribution modelling, or simply HOD modelling (see Cooray & Sheth 2002 for a review). Typically, central and satellite galaxies are treated separately, as each halo will contain one and only one of the former but may contain none or many of the latter (Kauffmann et al. 1999, Kravtsov et al. 2004, Zheng et al. 2005). The satellite galaxies assigned to a halo are usually assumed to be spherically distributed following a standard profile such as that of Navarro, Frenk, & White (1997). Attempts at including substructure or an environmental dependence have also been made (e.g. Giocoli et al. 2010, Gil-Marín, Jimenez & Verde 2011). Note that by assuming spherical symmetry some information is lost. As the simulations of Davis et al. (1985) first showed, cold dark matter haloes are typically strongly ellipsoidal. If the distribution of galaxies follows the mass distribution, this would leave an imprint on the galaxy correlation function on small scales. Furthermore, halo ellipticity may also have an effect on larger scales. If neighbouring haloes are aligned, as expected from tidal-torque theory, this will boost the correlation on scales corresponding to the typical separations between haloes.

The ellipticity and intrinsic alignment of dark matter haloes and their galaxy populations have been the subject of many earlier studies, e.g. Carter & Metcalfe (1980), Binggeli (1982), West (1989), Splinter et al. (1997), Jing & Suto (2002) and Bailin & Steinmetz (2005), and recently Paz et al. (2011) and Smargon et al. (2012). Most relevant to the current work are the studies by Smith & Watts (2005) and Zu et al. (2008). The former authors investigated the effects of halo triaxiality and alignment on the matter power spectrum in the halo model framework. Inspired by the results of simulations, they took a purely analytic approach in which they re-developed the halo model to account for ellipsoidal halo shapes. Zu et al. (2008), on the other hand, used the semi-analytical models of De Lucia & Blaizot (2007) to investigate environmental effects, including that of halo ellipticity, on the galaxy correlation function in both real and redshift space.

There exists a deeper connection between halo shape and clustering that we also explore in this paper. As Bett et al. (2007) and Faltenbacher & White (2010) have previously shown, at fixed mass the clustering of haloes depends on their shape. This is probably a reflection of assembly bias (i.e. the dependence of halo clustering on properties other than mass, Gao & White 2007). More specifically, the most spherical haloes in their samples cluster significantly more strongly than average, and the most aspherical more weakly. The known correlations between formation time and halo shape (e.g. Allgood et al. 2006, Ragone-Figueroa et al. 2010), and between formation time and clustering strength (e.g. Gao, Springel & White 2005, Wechsler et al. 2006, Wetzel et al. 2007, Jing, Suto & Mo 2007) point in this direction but may by themselves not

be strong enough to explain the magnitude of the effect. Here, we investigate whether this shape-dependence is also seen in the clustering of the galaxies. If so, this would bring us one step closer to measuring assembly bias directly in observations. We are herein also motivated by the results from Zhu et al. (2006) and Croton, Gao & White (2007), who showed that assembly bias in general is indeed expected to propagate to galaxy clustering.

In this paper, we expand upon previous work by investigating the effects of alignment and ellipticity on the galaxy correlation function using the Millennium Simulation (Springel et al. 2005) and the semi-analytic models of Guo et al. (2011). In Section 2, we discuss these simulations and our methods for quantifying the effects of alignment and ellipticity. We also outline our procedure for determining the shape-dependence of galaxy bias. We show our results in Section 3, and present our conclusions in Section 4.

## 2 METHODS

### 2.1 Simulation and SAM

We make use of the galaxy catalogues generated by Guo et al. (2011, hereafter G11), who implemented galaxy formation models on the Millennium Simulations (Springel et al. 2005, Boylan-Kolchin et al. 2009). The Millennium Simulation (MS) is a very large cosmological N-body simulation in which  $2160^3$  particles were traced from redshift 127 to the present day in a periodic box of side  $500 h^{-1}$  Mpc, comoving. The Millennium-II Simulation (MS-II) follows the same number of particles in a box of side  $100 h^{-1}$  Mpc and so has 125 times better mass resolution. Both simulations assume a  $\Lambda$ CDM cosmology with parameters based on a combined analysis of the 2dFGRS (Colless et al. 2001) and the first-year WMAP data (Spergel et al. 2003). These cosmological parameters, given by  $\{\Omega_m, \Omega_b, \Omega_\Lambda, \sigma_8, n_s, h\} = \{0.25, 0.045, 0.75, 0.9, 1.0, 0.73\}$ , are not consistent with the latest analyses of the CMB data, for example the seven-year WMAP results (Komatsu et al. 2011). In particular, the more recent data prefer lower  $\sigma_8$  and higher  $\Omega_m$  values.<sup>1</sup> We will only make relative comparisons between clustering statistics here, and do not expect our results to be significantly influenced by these small parameter differences (see Guo et al. 2012).

The galaxy formation models of G11 allow galaxies to grow at the potential minima of the evolving population of haloes and subhaloes in the simulations. Each Friends-of-Friends (FoF) group contains a central galaxy at the potential minimum of its main subhalo, and may contain many satellite galaxies at the centres of surrounding subhaloes. In some cases, due to tidal effects, a satellite galaxy may be stripped of its dark matter to the point where its subhalo is no longer identified as a bound substructure, turning the galaxy into an “orphan”. Such galaxies follow the orbit of the dark matter particle that had the highest binding energy immediately before subhalo disruption, except that their distance to the central galaxy is artificially decreased until they merge with it in order to mimic the effects of dynamical friction. We note that the treatment of

<sup>1</sup> See Angulo & White (2010) for a method to correct for this.

the orbits of orphans is approximate, and for example does not include the expected circularization of the orbits (e.g. Boylan-Kolchin, Ma & Quataert 2008). The models also include treatments of star formation, gas cooling, gas stripping, metal enrichment, supernova and AGN feedback, and galaxy mergers. For more details about the SAM, as well as the treatment of different types of galaxies, we refer to G11. For our purposes, it is enough to note that the predicted clustering of galaxies is quite a close match to that seen in the Sloan Digital Sky Survey (Guo et al. 2011).

## 2.2 Calculation of the galaxy correlation function

The galaxy two-point correlation function,  $\xi(r)$ , measures the clustering of galaxies as a function of scale. It effectively encodes the excess probability of finding a pair of galaxies at a given separation  $r$ , relative to the expectation for a uniform random distribution. In what follows, we will be interested in scales  $30 h^{-1} \text{ kpc} < r < 50 h^{-1} \text{ Mpc}$ , as these are both well-resolved and well-sampled by the simulation. In order to get accurate results over this full range, we calculate the correlation function by direct pair counts on small scales (i.e.  $r \lesssim 4 h^{-1} \text{ Mpc}$ ) and use an approximate but accurate method to calculate it on intermediate and large scales.

A direct calculation of this function scales as the number of galaxies squared and is thus unfeasible for the large sample analysed here. We therefore speed up the calculation by mapping galaxies onto a grid, and we correlate the mean density contrast in each grid cell with that of every other (a method previously employed by, for example, Barriga & Gaztañaga 2002, Eriksen et al. 2004 and Sánchez, Baugh & Angulo 2008). We improved the performance on intermediate scales by folding the density field onto itself before its autocorrelation is calculated (see e.g. Jenkins et al. 1998). We do not go into these methods here, but note that tests against higher-accuracy calculations show that the error in the ratio of the correlation functions, which is the relevant quantity for our main results, is less than 1 per cent on all scales considered. For the correlation functions we calculate to determine the galaxy bias a direct pair count over the full range of scales is feasible, as there we only consider relatively small subsets of galaxies (see §3.2).

## 2.3 Testing the importance of alignment and ellipticity

During their lifetime, haloes merge and may accrete more subhaloes. The accretion of mass is not isotropic since matter flows in preferentially along filaments (see e.g. Tormen, Bouchet & White 1997, Colberg et al. 1999 or more recently Vera-Ciro et al. 2011). As a result, the distribution of subhaloes and thus galaxies within a FoF group is generally not isotropic either, but is instead approximately ellipsoidal, following the mass and aligning with surrounding large-scale structure (see e.g. Angulo et al. 2009). To test whether alignment with neighbouring structure has an effect on clustering statistics, we randomly rotate the haloes around their centres and see if this systematically alters the galaxy correlation function. More precisely, we rotate the satellite population of each FoF group bodily around the central galaxy to a new randomly chosen orientation, and

we repeat this process for every FoF group in the simulation. We stress that this transformation preserves the numbers, properties, and relative positions of the galaxies in every halo; only the orientations of the distributions change. We then calculate the galaxy correlation function for the new distribution, and compare it to the original. If alignment with large-scale structure is important, one would expect to see the correlation decrease systematically on scales slightly larger than individual haloes. To estimate the uncertainty in our results, we have repeated this process 25 times, each time with a different set of randomly chosen angles.

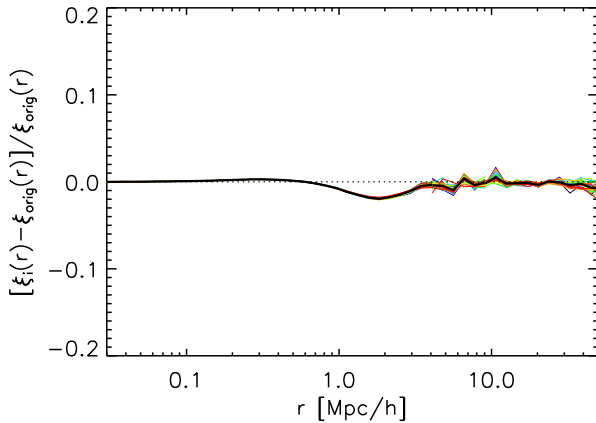
The effect of halo ellipticity is tested in a similar way. Here, we randomly rotate the position of each individual satellite galaxy around its central, rather than rotating all satellites together. In this way, the galaxy distribution within each halo is sphericalized. Since the distribution of galaxies within haloes is typically ellipsoidal, this process should increase the average distance between galaxies, thus decreasing the correlations between galaxies in the same halo. We note that Zu et al. (2008) investigated the effect of halo ellipticity in the same way.

## 2.4 Testing the dependence of galaxy bias on halo shape

Faltenbacher & White (2010) showed that the clustering of haloes depends on the shape of the halo, defined as  $s = c/a$ , where  $c$  and  $a$  are eigenvalues of the inertia tensor ( $a > b > c$ ). Specifically, they showed that the large-scale bias of haloes with more spherical shapes is larger than average, while the inverse is true for the most aspherical haloes. They also found that this difference decreases with equivalent peak height,  $\nu(M, z) = \delta_c(z)/\sigma(M, z)$ , where  $\sigma(M, z)$  is the root-mean-square linear overdensity within a sphere which contains the mass  $M$  in the mean, and  $\delta_c(z)$  is the linear overdensity threshold for collapse at redshift  $z$ . Here, we are interested in seeing if this shape-dependent clustering, which might reflect the assembly bias of the haloes, is also recovered from the galaxy distribution.

We use the halo shape data from Faltenbacher & White (2010), who calculated the inertia tensor from the dark matter particles belonging to the most-pronounced subhalo<sup>2</sup> of each FoF halo, which on average comprises  $\sim 80$  per cent of its mass. To ensure that the shapes were accurately determined, only haloes with at least 700 particles were considered, corresponding to a minimum (sub)halo mass  $M = 6.02 \times 10^{11} h^{-1} \text{ M}_\odot$ . We compare this to the shape measured from the galaxy distribution in the same way, using all galaxies with stellar masses  $M_* > 10^9 h^{-1} \text{ M}_\odot$  within a sphere of radius  $R_{200}$ , defined as the radius enclosing 200 times the mean density of the Universe, centred on the central galaxy. Using such a distance cut makes it easier to compare our results to observations – in fact, a similar procedure is often followed when determining the richness of

<sup>2</sup> The most-pronounced subhalo may differ from the most massive subhalo only if the FoF groups hosts two or more subhaloes with roughly the same mass. In this case one of these is arbitrarily assigned to be the most massive. The most-pronounced subhalo is more consistently defined by using the halo merger tree. It is also the subhalo hosting the most luminous galaxy in the FoF group.



**Figure 1.** The effect of halo alignment on the galaxy correlation function. The x-axis shows the real-space separation  $r$ , while the y-axis shows the fractional difference between the correlation function after random bodily rotations are applied, and that of the original, unrotated sample. All simulated galaxies with  $M_* > 10^9 h^{-1} M_\odot$  from the  $z = 0$  catalogue of G11 have been used here. The bin size is roughly 0.07 dex. Each of the 25 thin, coloured lines represents a different set of random rotations, and the thick, black line shows the average of these. There is a clear signal around  $r \approx 1.8 h^{-1} \text{Mpc}$ , where the correlation function is lowered by roughly 2 per cent.

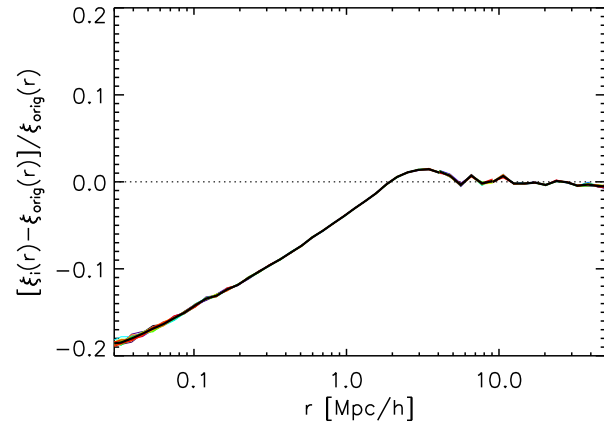
real groups and clusters. Rejecting galaxies outside the virial radius slightly biases us to measure more spherical shapes, but we have checked that this effect is small and does not significantly affect our results.

After splitting the galaxies by the shape of their halo (measured either from the dark matter or from the galaxies themselves) we determine the large-scale galaxy bias factor  $b_{\text{gal}}$  for each subsample. Here, too, we follow Faltenbacher & White (2010), who in turn followed the approach of Gao & White (2007). The bias is computed as the relative normalization factor that minimizes the mean square of the difference  $\log(\xi_{\text{gm}}) - \log(b_{\text{gal}}\xi_{\text{mm}})$  for four bins spaced equally in  $\log r$  in the range  $6 < r < 20 h^{-1} \text{Mpc}$ . Here  $\xi_{\text{mm}}$  is the dark matter autocorrelation function and  $\xi_{\text{gm}}$  is the cross-correlation function of galaxies and dark matter. Note that unlike Faltenbacher & White (2010) we are only interested in the results at  $z = 0$ .

### 3 RESULTS

#### 3.1 Alignment and ellipticity

We will first discuss our results for “bodily” rotations, which test the effect of halo alignment. Figure 1 shows the fractional difference between the correlation functions of the “rotated” and original samples, plotted against the real-space separation  $r$ . We have only used those galaxies from the catalogue generated by G11 that have a stellar mass  $M_* > 10^9 h^{-1} M_\odot$ , as the Millennium Simulation is not complete below this limit. This provides a sample of 5 200 801 galaxies. We note that increasing this mass limit by a factor of ten does not influence our results, either qualitatively or quantitatively. All random rotations are applied prior to the mass cut in order to avoid problems in cases

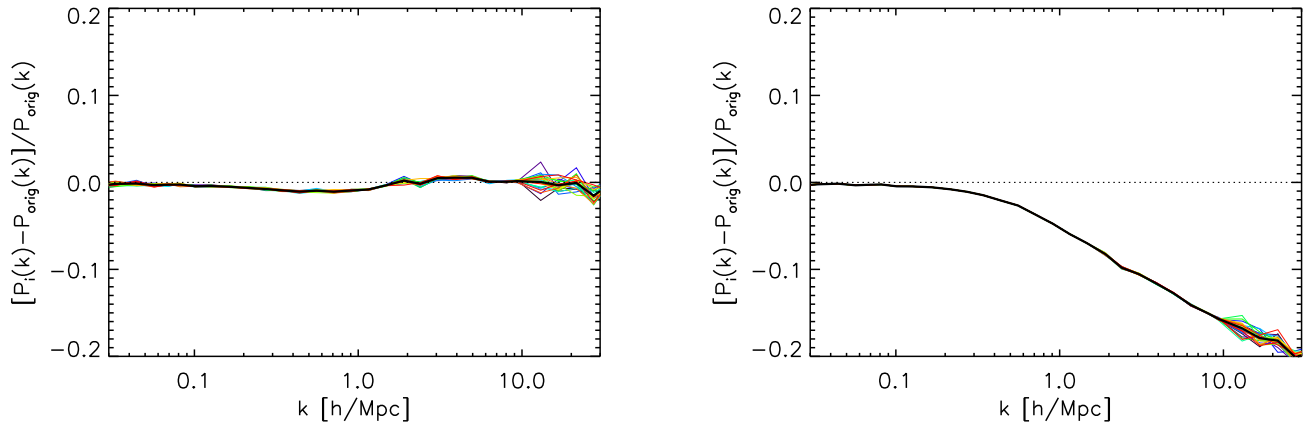


**Figure 2.** The effect of halo ellipticity on the galaxy correlation function. The bin size, axes and lines are as in Figure 1, although the bodily rotations have been replaced by random independent rotations of satellites around their central galaxies. A peak of 1–2 per cent can be seen around  $r \approx 3.5 h^{-1} \text{Mpc}$ , but the largest effect is seen on small scales, where the correlation function is systematically lowered by up to  $\sim 20$  per cent. Note also that the scatter has been greatly reduced relative to Figure 1. This is mainly due to the larger number of random rotations used when rotating satellites separately.

where a central galaxy below the limiting mass has satellites above it. Coloured lines indicate different sets of rotations, while the thick, black line shows the average of these. There is clearly a significant dip around  $r \approx 2 h^{-1} \text{Mpc}$ , with a depth of 2 per cent. This is due to the disruption of the alignment between haloes and surrounding structure. Note that the scatter is extremely low, due to the large number of objects (in fact, the uncertainty at large scales is dominated by the errors due to our approximate calculation of the correlation). Neglecting the orientation of haloes when populating them with galaxies will therefore have a modest, but significant, effect on the derived correlation function.

The ellipsoidal shape of the haloes, and the fact that the galaxy distribution follows this shape, is a more significant factor when modelling the galaxy distribution. Figure 2 shows the result of applying independent rotations, which sphericalize the galaxy distributions within haloes. This substantially suppresses correlations for  $r \lesssim 2 h^{-1} \text{Mpc}$ , with a  $\sim 20$  per cent effect on the smallest scales probed here. The ellipsoidal shape of the galaxy distribution within haloes significantly reduces the typical separations of pairs within them. This is compensated by a 1–2 per cent stronger correlation around  $r \approx 3.5 h^{-1} \text{Mpc}$ . We conclude that models that assume spherical profiles for the distribution of galaxies within haloes will underestimate the galaxy correlation function by up to  $\sim 20$  per cent, depending on the smallest scale considered. Note that the scatter is even smaller than before on all scales, which is due to the increased number of degrees of freedom here. This result is in excellent agreement with that of Zu et al. (2008), who applied similar methods to all galaxies from De Lucia & Blaizot (2007) with  $r$ -band luminosities  $M_r < -19$ .

One might worry that the necessarily artificial treatment of orphan, or “type 2”, galaxies in the galaxy formation models of G11 influences these results. G11 already showed that the inclusion of the orphans is critical if the radial dis-



**Figure 3.** Same as Figures 1 and 2, but now for the fractional differences in the galaxy power spectrum versus the wave number  $k$ . *Left:* Result when testing for alignment. Again a weak but systematic signal of a few per cent can be seen, now between  $k \approx 0.1 h \text{ Mpc}^{-1}$  and  $k \approx 2 h \text{ Mpc}^{-1}$ . *Right:* Result when testing for ellipticity. A monotonic decline in power sets in at  $k \approx 0.1 h \text{ Mpc}^{-1}$ , reaching roughly 20 per cent at  $k = 30 h \text{ Mpc}^{-1}$ , matching the result found in Figure 2. This again demonstrates the importance of taking the ellipsoidal shape of the galaxy distribution within haloes into account.

tribution of galaxies within rich clusters in the Millennium Simulation is to agree both with observations and with the much higher-resolution MS-II. We have investigated effects on our analysis by repeating it with these galaxies removed, reducing our sample size by  $\sim 24$  per cent and spoiling the relatively good agreement of its small-scale correlation with observation. This removal significantly amplifies the signal found for the effects of alignment. This is because the orphan galaxies are primarily located near halo centres. Once they are removed, galaxies that do contribute to the alignment signal receive more weight. Orphan removal also changes the signal found for the effects of ellipticity, modestly boosting it down to  $r \approx 0.1 h^{-1} \text{ Mpc}$ . As a further check that the distribution of orphans in the simulation is realistic, we have examined the shapes of the galaxy distributions of massive haloes (specifically,  $14 < \log_{10}(M_{\text{FoF}}/[h^{-1} \text{ M}_{\odot}]) < 14.5$ ) in the MS and MS-II, again using G11's galaxy catalogues and considering only galaxies with stellar masses  $M_* > 10^9 h^{-1} \text{ M}_{\odot}$ . The better mass resolution of the MS-II results in far fewer orphans in this mass range, and consequently the positions of galaxies in MS-II are determined more accurately. Nevertheless, the shapes of the galaxy distributions agree very well, thus implying that the distribution of orphans in the MS is consistent with the distribution of similar, but unstripped, galaxies in the MS-II. We also found that these shapes agree very well with those of the dark matter haloes themselves (see Bett et al. 2007). A more detailed discussion of the shapes is beyond the scope of this paper.

For completeness, we have also compared the galaxy-galaxy power spectra of the rotated and unrotated samples. The results are shown in Figure 3. Three foldings in total were used to calculate the power spectra over the full range shown, each with a fold factor of six (i.e. each folding maps the particle distribution to 1/216th of the volume). The power spectra were re-binned logarithmically to resemble the bins used for the galaxy correlation functions, and to reduce noise. Shot noise, which dominates the power for  $k \gtrsim 10 h \text{ Mpc}^{-1}$ , was subtracted. The left-hand figure shows the fractional differences that result from applying bodily ro-

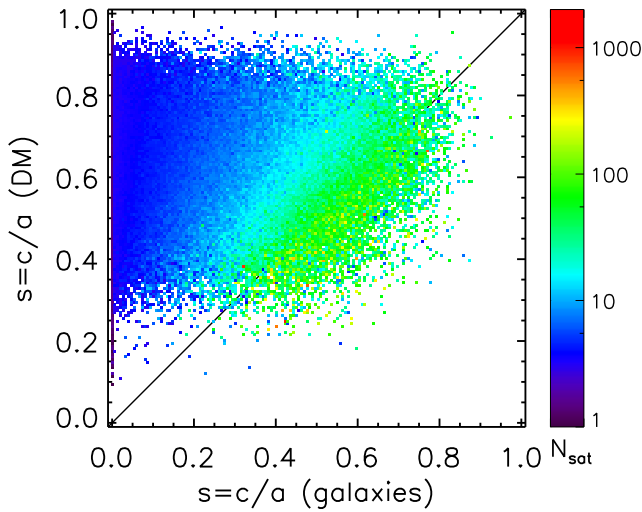
tations. Just as for the correlation function, there is a weak but clear dip of 1–2 per cent present which reflects the alignment of haloes with surrounding structure. As expected, the right-hand figure shows a much stronger decrease in power, up to 20 per cent on the smallest scales considered. This again demonstrates the importance of taking the ellipsoidal distribution of galaxies within FoF groups into account in, for example, models that use that use the full shape of the power spectrum to extract cosmological parameters.

Our results differ from those found by Smith & Watts (2005). Like ourselves, they find that the scale at which the contribution from alignments to the power spectrum is maximal is  $\sim 0.5 h \text{ Mpc}^{-1}$ , but they show that the relative contribution of alignments is strongly model-dependent, varying from  $10^{-12}$  to 10 per cent. Furthermore, they find that when haloes are assumed to be spherical, the power is higher than when they are ellipsoidal by up to 5 per cent. Not only is the effect we find significantly stronger and increasing towards smaller scales up to at least  $k = 30 h \text{ Mpc}^{-1}$ , but its sign is opposite. We attribute these differences to the fact that in both models explored by Smith & Watts (2005), the radially averaged density profiles are not conserved when transforming the haloes from spherical to triaxial, making a comparison with our own results difficult.

### 3.2 Shape-dependent galaxy bias

Having established that the shape of the galaxy distribution significantly affects the small-scale clustering, we now investigate how the shape-dependent assembly bias affects clustering. We first examine how well the shape measured from the galaxy distribution corresponds to that measured from the dark matter particles belonging to the most-pronounced substructure. Note that while the galaxies trace the shape of the FoF halo very well, this is not necessarily the case for its most-pronounced substructure. Additionally, even if the galaxy distribution traces the dark matter perfectly, the right shape may still not be recovered if only a small number of galaxies is available to sample the halo.

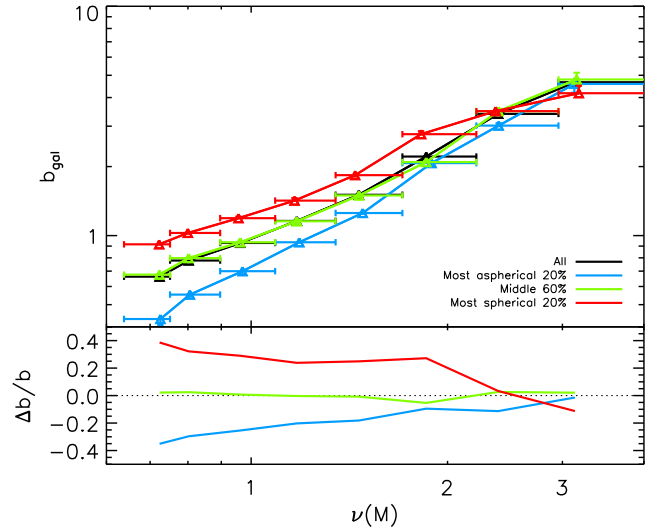
The results of this shape comparison are presented in



**Figure 4.** A comparison of the halo shape measured from the dark matter and that measured from the galaxy distribution. Here shape is defined as the ratio of the smallest to the largest eigenvalue of the inertia tensor. The shape on the horizontal axis is computed using all galaxies satisfying  $M_* > 10^9 h^{-1} M_\odot$  within  $R_{200}$ . For the shape on the vertical axis, only CDM particles belonging to the most-pronounced substructure are included, and only if the number of particles is at least 700. Pixels are colour-coded by mean number of galaxies per halo. The “true” shape is recovered more accurately when the number of galaxies increases; some scatter remains however, as the galaxy distribution does not perfectly trace the main subhalo.

Figure 4. Here the shape measured from the galaxy distribution is shown on the horizontal axis, while the shape measured from the dark matter is on the vertical axis. Each pixel is colour-coded by the mean number of satellites,  $N_{\text{sat}}$ , satisfying  $M_* < 10^9 h^{-1} M_\odot$  and  $|r_{\text{sat}} - r_{\text{cen}}| < R_{200}$ . A low value of  $s = c/a$  indicates that the halo is (measured to be) very aspherical, while a perfectly spherical halo would have  $s = 1$ . It is immediately clear that a low number of satellites leads to a severe underestimate of  $s$ . This is expected: together with the central galaxy, any two satellites will define a plane, ensuring that  $c = 0$ . It is only when the halo is sampled by a large enough ensemble of points that the inertia tensor can be determined accurately. Figure 4 illustrates that one needs  $N_{\text{sat}} \gtrsim 30$  to get an unbiased and accurate shape estimate. However, there is always a significant amount of scatter around the diagonal. This is due to the galaxies tracing the mass, i.e. the shape of the whole FoF group, and not just that of the most-pronounced substructure.

Next, we split our galaxy sample by the shape measured from the halo dark matter distribution and calculate the galaxy bias factor as function of equivalent peak height,  $b_{\text{gal}}(\nu)$ , following the method described in §2.4. As we can only use galaxies for which a dark matter shape has been determined – i.e. those in FoF haloes of which the most massive substructure is comprised of at least 700 particles – our galaxy sample is reduced to 2953050 galaxies. The results are shown in Figure 5. Here the upper panel shows the galaxy bias determined for each sample at a given equivalent peak height. The points show the median value, horizontal

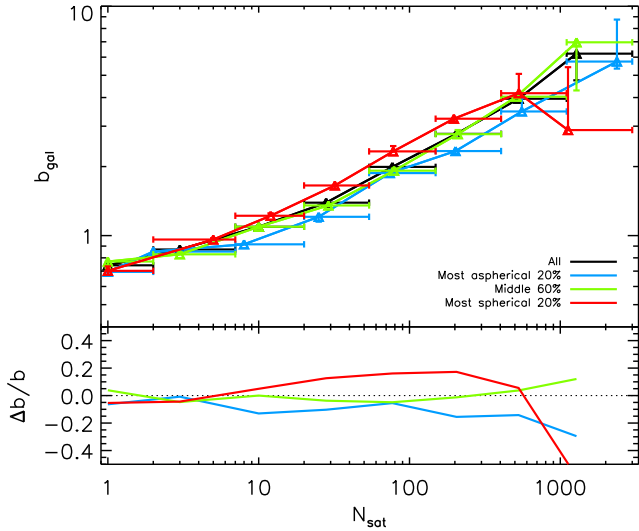


**Figure 5.** The dependence of galaxy bias,  $b_{\text{gal}}$ , on shape, as a function of peak height at  $z = 0$ . The black line in the top panel shows the galaxy bias of the full galaxy sample, while coloured lines show the bias of subsamples split by halo shape measured from the dark matter distribution. In the bottom panel the fractional differences of the bias of these subsamples relative to the full sample are shown. The vertical error bars show  $1\sigma$  deviations calculated from 50 bootstrap resamplings of the galaxy catalogues. We find that galaxies in the most spherical (aspherical) haloes are strongly biased (antibased) relative to the full sample. The difference can be as much as 40 per cent for  $\nu \approx 0.7$ . For  $\nu \gtrsim 2.5$  statistical uncertainties, due to the low number of high-mass haloes, begin to play an important role.

error bars indicate the width of the bin, and vertical error bars show  $1\sigma$  deviations calculated from 50 bootstrap resamplings of the galaxy catalogues. The black line shows  $b_{\text{gal}}(\nu)$  for the full sample of galaxies, while the coloured lines show the bias for the different subsamples. In the bottom panel, the fractional difference between these subsamples and the full sample is shown. It is immediately clear that there is a strong dependence on shape: at the lowest equivalent peak heights probed here, the bias of the galaxies in the most spherical (aspherical) 20 per cent of haloes is up to 40 per cent higher (lower) than that of the full sample. This shows that the shape-dependence of the halo bias found by Bett et al. (2007) and Faltenbacher & White (2010) is also strongly present in the clustering of the galaxies. The effect grows weaker with increasing  $\nu$ . However, for  $\nu \gtrsim 2.5$  statistical uncertainties begin to dominate, due to the low number of haloes available at high equivalent peak heights. Note that this is not seen in the data of Faltenbacher & White (2010) as they combine the data from different redshifts, while we only consider  $z = 0$ .

We then repeat this exercise, but this time we split our galaxy sample by the shape measured from the distribution of the galaxies themselves. However, we expect low-mass haloes to host only a few satellite galaxies, leading to very unreliable estimates of the halo shape (see Figure 4). In order to separate the signal we are looking for – i.e. the shape-dependence of the galaxy bias – from the unwanted bias introduced by using too few galaxies in the shape measurement, we now consider  $b_{\text{gal}}$  as a function of the number





**Figure 6.** As Figure 5, but now showing the bias as a function of the number of satellite galaxies and split by the halo shape measured from galaxies within  $R_{200}$ . At low  $N_{\text{sat}}$  no significant shape-dependence of  $b_{\text{gal}}$  is recovered, due to the extremely unreliable shape determinations that follow from using only a handful of galaxies to sample the halo. For  $10 \lesssim N_{\text{sat}} \lesssim 400$ , however, we find again that galaxies in more spherical (aspherical) haloes have a significantly higher (lower) bias than average. At higher  $N_{\text{sat}}$  our results are again dominated by poor statistics.

of satellites per halo,  $N_{\text{sat}}$ , instead of the equivalent peak height  $\nu$ . An additional advantage of this approach is that  $N_{\text{sat}}$  is directly observable. We note, however, that we obtain almost identical results when considering galaxy bias as a function of peak height instead of  $N_{\text{sat}}$ . Since we do not use the dark matter particle data in this case, we are no longer constrained by needing haloes with at least 700 particles. However, as we can now only consider haloes with at least one satellite galaxy that satisfies both  $M_* > 10^9 h^{-1} M_\odot$  and  $|r_{\text{sat}} - r_{\text{cen}}| < R_{200}$ , we are left with a sample of 2 566 441 galaxies. The results are shown in Figure 6. At the lowest value of  $N_{\text{sat}}$ , no significant effect can be seen. But as  $N_{\text{sat}}$  grows, increasing the accuracy of the shape determinations, we again see a clear dependence of  $b_{\text{gal}}$  on the shape  $s$ : galaxies in more spherical haloes are  $\sim 20$  per cent more strongly clustered than average. The inverse is true for galaxies in the most aspherical haloes. When  $N_{\text{sat}}$  grows too high our results are once more dominated by statistical errors, due to the low number of high-mass haloes (hosting at least several hundreds of satellites) available.

These results show that halo assembly bias in the form of a shape-dependent clustering strength propagates to the clustering of galaxies, and can therefore in principle be measured in sufficiently large surveys. In order to carry out such a task, a large galaxy survey with appropriately defined group catalogues is needed.

#### 4 SUMMARY

We have investigated the effects of halo alignment with larger scale structure and of halo ellipticity on galaxy correlation functions, using the Millennium Simulations

(Springel et al. 2005, Boylan-Kolchin et al. 2009) and the galaxy formation models of Guo et al. (2011). By rotating satellite galaxies in FoF groups around their central galaxies, either coherently for each halo or independently for each satellite, and then comparing the correlation function of the resulting galaxy distribution to the original one, we were able to quantify the importance of taking halo alignment and non-sphericity into account. Furthermore, by measuring the shape of the haloes as traced by the galaxies we were able to investigate the propagation of shape-dependent assembly bias to the clustering of galaxies. Only galaxies with stellar masses  $M_* > 10^9 h^{-1} M_\odot$  were considered in our analysis, though we note that increasing this mass limit by a factor of ten does not influence our results. Our findings can be summarized as follows:

- The effects on the galaxy correlation function of the alignment of haloes with larger-scale structure are small. The main effect of disrupting this alignment is a 2 per cent reduction in correlation amplitude around  $r \approx 1.8 h^{-1} \text{Mpc}$ , with minor effects of at most 1 per cent at smaller scales.
- The ellipsoidal shapes of the galaxy distributions within individual haloes have a much stronger influence on galaxy correlations. By sphericalizing these galaxy distributions (i.e. randomizing the angular positions of satellites while keeping the distance from the central galaxy fixed), the correlation function is raised by up to 2 per cent around  $r \approx 3.5 h^{-1} \text{Mpc}$ , but greatly reduced for  $r \lesssim 1.5 h^{-1} \text{Mpc}$ , by up to  $\sim 20$  per cent on the smallest scale probed,  $r = 30 h^{-1} \text{kpc}$ . This confirms the results of Zu et al. (2008). The effect on the galaxy power spectrum extends to scales as large as  $k = 0.1 h \text{Mpc}^{-1}$ .
- The assembly bias of haloes, as characterized by the dependence of clustering on halo shape, is reflected in the clustering of galaxies. The effect is strongest at low equivalent peak heights: at  $\nu \approx 0.7$ , the galaxy bias of galaxies in the 20 per cent most spherical and most aspherical haloes deviate from the average by 40 per cent.
- Even if the shape of the halo cannot be measured directly, but is instead estimated from galaxies within one virial radius of the central galaxy, the effect of assembly bias is clearly visible.

By using the plane-parallel approximation and ignoring evolution, we have checked that comparable results are obtained for the effects of alignment and ellipticity on the redshift-space correlation functions. Models that assume a spherically symmetric profile for the galaxy distribution, such as HOD models, will therefore significantly underestimate galaxy correlations and power spectra on sub-Mpc scales.

Furthermore, we have demonstrated that the shapes of haloes and of the galaxy distributions within them can be strongly correlated with their clustering. This effect should be measurable in galaxy redshift surveys. With the help of extensive group catalogues it should therefore be possible to measure assembly bias directly.

#### ACKNOWLEDGMENTS

The authors thank Andreas Faltenbacher for kindly providing dark matter shape determinations for haloes in the Millennium Simulation. The Millennium Simulation databases

used in this paper and the web application providing on-line access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory. This work was supported by the Marie Curie Initial Training Network CosmoComp (PITN-GA-2009-238356) and by Advanced Grant 246797 "GALFORMOD" from the European Research Council.

## REFERENCES

Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, *MNRAS*, 367, 1781

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, *MNRAS*, 399, 983

Angulo R. E., White S. D. M., 2010, *MNRAS*, 405, 143

Bailin J., Steinmetz M., 2005, *ApJ*, 627, 647

Barriga J., Gaztañaga E., 2002, *MNRAS*, 333, 443

Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 3101

Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, *MNRAS*, 376, 215

Binggeli B., 1982, *A&A*, 107, 338

Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, *MNRAS*, 383, 93

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150

Carter D., Metcalfe N., 1980, *MNRAS*, 191, 325

Colberg J. M., White S. D. M., Jenkins A., Pearce F. R., 1999, *MNRAS*, 308, 593

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781

Colless M. et al., 2001, *MNRAS*, 328, 1039

Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1

Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371

De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2

Eriksen H. K., Lilje P. B., Banday A. J., Górski K. M., 2004, *ApJS*, 151, 1

Faltenbacher A., White S. D. M., 2010, *ApJ*, 708, 469

Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66

Gao L., White S. D. M., 2007, *MNRAS*, 377, L5

Gil-Marín H., Jimenez R., Verde L., 2011, *MNRAS*, 414, 1207

Giocoli C., Bartelmann M., Sheth R. K., Cacciato M., 2010, *MNRAS*, 408, 300

Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2012, preprint (arXiv:1206.0052)

Guo Q. et al., 2011, *MNRAS*, 413, 101

Jenkins A. et al., 1998, *ApJ*, 499, 20

Jing Y. P., Suto Y., 2002, *ApJ*, 574, 538

Jing Y. P., Suto Y., Mo H. J., 2007, *ApJ*, 657, 664

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, *MNRAS*, 303, 188

Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201

Komatsu E. et al., 2011, *ApJS*, 192, 18

Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35

Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493

Paz D. J., Sgró M. A., Merchán M., Padilla N., 2011, *MNRAS*, 414, 2029

Ragone-Figueroa C., Plionis M., Merchán M., Gottlöber S., Yepes G., 2010, *MNRAS*, 407, 581

Sánchez A. G., Baugh C. M., Angulo R., 2008, *MNRAS*, 390, 1470

Smargon A., Mandelbaum R., Bahcall N., Niederste-Ostholt M., 2012, *MNRAS*, 423, 856

Smith R. E., Watts P. I. R., 2005, *MNRAS*, 360, 203

Spergel D. N. et al., 2003, *ApJS*, 148, 175

Splinter R. J., Melott A. L., Linn A. M., Buck C., Tinker J., 1997, *ApJ*, 479, 632

Springel V., 2010, *ARA&A*, 48, 391

Springel V. et al., 2005, *Nature*, 435, 629

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726

Tormen G., Bouchet F. R., White S. D. M., 1997, *MNRAS*, 286, 865

Vera-Ciro C. A., Sales L. V., Helmi A., Frenk C. S., Navarro J. F., Springel V., Vogelsberger M., White S. D. M., 2011, *MNRAS*, 416, 1377

Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71

West M. J., 1989, *ApJ*, 347, 610

Wetzel A. R., Cohn J. D., White M., Holz D. E., Warren M. S., 2007, *ApJ*, 656, 139

White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52

Zheng Z. et al., 2005, *ApJ*, 633, 791

Zhu G., Zheng Z., Lin W. P., Jing Y. P., Kang X., Gao L., 2006, *ApJ*, 639, L5

Zu Y., Zheng Z., Zhu G., Jing Y. P., 2008, *ApJ*, 686, 41

This paper has been typeset from a  $\mathrm{T}_{\mathrm{E}}\mathrm{X}/\mathrm{L}^{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$  file prepared by the author.